

Approximate and Estimated Saddlepoint Approximations

Pamela A. Ohman*
University of Florida

George Casella†
Cornell University

August 21, 1998

Keywords: asymptotics, small sample statistics, bootstrap, Edgeworth expansions

Abstract

Saddlepoint approximations are highly accurate tools for approximating the density or distribution of a statistic, in particular a statistic that is found using an estimating equation. In the classical theory of saddlepoint approximations, it is assumed that the cumulant generating function (*cgf*) of the underlying distribution is known, allowing approximations to achieve a relative error of order $O(n^{-1})$. This translates into excellent accuracy even for small sample sizes. However, for many reasons (simplicity, practicality) the cumulant generating function may be approximated with a truncated Taylor series. Moreover, in practice it may also be necessary to estimate the parameters of this approximation of the *cgf* before utilizing the saddlepoint formula. We show that saddlepoint approximations for the cumulative distribution of the quantity $n^{1/2}(\bar{X} - \mu)$ obtain a relative accuracy to the true distribution of $O(n^{-1/2})$. In the case where the parameters of the truncated Taylor series are estimated with order of accuracy $\delta > 0$, the saddlepoint approximation maintains $O(n^{-\delta})$ accuracy. Simulation results for small samples are also presented.

*Supported in part by National Science Foundation Training Grant DMS-956682.

†Supported by National Science Foundation Grant DMS-9625440. This is technical report BU-1418-M in the Department of Biometrics, Cornell University. File estsad3.tex.

1 Introduction

Saddlepoint approximations were introduced for use in statistics by Daniels (1954, 1980, 1983). These approximations have been applied primarily in the parametric setting, originally to approximate the density or distribution of the univariate sample mean. More recently, they have been applied in more general settings to find densities or distributions of permutation distributions (Robinson 1982), M-estimators and functions of M-estimators (Gatto and Ronchetti, 1996), and L-statistics (Easton and Ronchetti, 1986). Book length treatments of saddlepoint and associated expansions are given by Field and Ronchetti (1990), Hall (1992), Kolassa (1994), and Jensen (1995). The review articles of Reid (1988, 1991), and the more elementary introduction by Goutis and Casella (1998) also provide entries to this topic.

In the case of the univariate sample mean, the strength of the saddlepoint approximation lies in the fact that the relative error is uniformly bounded and is often $O(n^{-1})$ (Reid, 1988; Daniels, 1987). Thus even in the tails of the distribution, this approximation is very accurate. This is in contrast to the Edgeworth series approximations, which have an absolute instead of relative error, and whose relative error blows up in the tails of the distribution.

A major drawback is that the use of the full saddlepoint approximation requires the determination of the cumulant generating function of the statistic of interest which, for complex statistics, may not be simple (or even possible). However, determining the first four moments is often possible. In these cases, a saddlepoint approximation of the distributions of these statistics can itself be approximated by using a Taylor series approximation of the cumulant generating function (typically using the first four terms). This approach was proposed by Easton and Ronchetti (1986) and may be used in a wide range of situations, from the fully parametric to the fully non-parametric. The approach of Easton and Ronchetti was modified by Wang (1992) to ensure that the estimated *cgf* has a unique minimum for use in the saddlepoint approximation. The modified *cgf* is a weighted average of the Easton and Ronchetti truncated *cgf* and an approximate normal *cgf*. However, little work has been published which analytically demonstrates the accuracy of these approximate saddlepoint formulas. An exception is Feuerverger (1989), who has analytically studied the accuracy of a similar approximate saddlepoint approximation, that which results from substituting the empirical *cgf* for the true *cgf* in the usual saddlepoint formula for the density of the mean. This so-called *empirical* saddlepoint approximation was originally proposed by Davison and Hinkley (1988), as a way to approx-

imate the bootstrap distribution. While much of this previous work focuses on the saddlepoint approximation to the density, we focus on saddlepoint approximations to the cumulative distribution, which is directly useful for inference via confidence intervals and hypothesis tests.

By taking an approach similar to Feuerverger (1989), which requires studying the saddlepoint approximation of $n^{1/2}(\bar{X} - \mu)$, we address the following issues:

- (i). The accuracy of the *cgf* approximation using four known cumulants, the *approximate* saddlepoint approximation;
- (ii). The accuracy of the *cgf* approximation using the empirical *cgf*, the *estimated nonparametric* saddlepoint approximation;
- (iii). The accuracy of the *cgf* approximation using four estimated cumulants (parameter estimates within the parametric *cgf*), the *estimated parametric* saddlepoint approximation.

We demonstrate analytically that no accuracy is lost in the saddlepoint formula when using the Easton-Ronchetti method (Case *i*), given the saddlepoint exists. Furthermore, we extend this approach to the empirical saddlepoints in order to produce estimated distributions which allow for inference about the mean $\mu = E(X)$ (Case *ii*). In addition, similar proofs lead to understanding the effect of estimating parameters on the final saddlepoint approximation in the parametric setting.

The basic saddlepoint approximation is given in Section ?? . Theory for the cases described above is given in Section ?? . To demonstrate application of the theory, in Section ?? , a simulation study is performed. The last section provides some final conclusions and directions for further research.

2 The Saddlepoint Approximation

Suppose that the independent random variables X_i , $i = 1, \dots, n$, are drawn from a density $f(x)$. For a single observation $X \sim f(x)$, denote its moment and cumulant generating functions by $M(t) = E(e^{tX})$ and $K(t) = \log M(t)$.

From these observations, inference based on the mean \bar{X} (here taken to be of dimension 1) is of interest. The saddlepoint approximation to this distribution can be formulated in two ways. The first is a Laplacian approximation of the integrated saddlepoint density (Robinson 1982, Daniels 1983, 1987), both of which approximate direct numerical integration of the

saddlepoint density. The other is a direct saddlepoint approximation of the tail area, often called the Lugannani and Rice (1980) formula. Daniels (1987) shows these two approaches to be comparable, and they work especially well, in the tails of the distribution. This is of particular importance for hypothesis tests and confidence intervals. We focus on the former formula because of its somewhat simpler form. However, the same results follow for both approximations due to their asymptotic equivalence.

Thus, the basic integrated saddlepoint formula for the cumulative distribution of \bar{X} is given by

$$\Pr(\bar{X} > x) = \frac{\exp\{n(K(t) - tx)\}}{t(2\pi nK''(t))^{1/2}} \left\{ 1 + \frac{1}{n} \left[\frac{1}{8}\lambda_4 - \frac{5}{24}\lambda_3^2 - \frac{\lambda_3}{2t(K''(t))^{1/2}} - \frac{1}{t^2K''(t)} \right] + O(n^{-2}) \right\}$$

where t is the solution to $K'(t) = x$ and for $r = 1, \dots, 4$, $\lambda_r = K^{(r)}(t)/\{K''(t)\}^{r/2}$ with $K^{(r)}$ the r th derivative of the *cgf*. Note that t is a function of x as well as any parameters on which the cumulant generating function depends.

Transforming this to find the distribution of $V = n^\delta(\bar{X} - \mu)$, for $\delta \leq 1/2$ creates no loss in accuracy. The parameter $\mu = E(X)$ is given by the usual first cumulant $K'(0)$. The distribution for the transformed variable is given by

$$\begin{aligned} \Pr(V > v) &= \Pr(\bar{X} > n^{-\delta}v + \mu) \\ &= \frac{\exp\{n(K(t) - n^{-\delta}tv - t\mu)\}}{t(2\pi nK''(t))^{1/2}} \\ (1) \quad &\times \left\{ 1 + \frac{1}{n} \left[\frac{1}{8}\lambda_4 - \frac{5}{24}\lambda_3^2 - \frac{\lambda_3}{2t(K''(t))^{1/2}} - \frac{1}{t^2K''(t)} \right] + O(n^{-2}) \right\} \end{aligned}$$

where, now, t is the solution to $K'(t) = n^{-\delta}v + \mu$.

3 Implications of Estimating the *cgf*

If the cumulant generating function is unknown, even if only in part, then approximations to or estimations of the *cgf* must be made before applying the saddlepoint formula. In this section, we first give a theoretical explanation of the effect of approximation or estimation on the resulting saddlepoint approximation. We describe in detail the various approximations, parametric and nonparametric, and the implications of the theory for these approximations.

3.1 General Theory - Truncated Taylor Series of the *cgf*

If it is the case that the cumulant generating function is known only up to a finite numbered cumulant, or that the entire cumulant generating function is intractable to work with in the saddlepoint formula, the use of an approximate cumulant generating function must be considered. In particular, approximate the sampling *cgf* using the first k cumulants, assuming all other cumulants are zero. Thus

$$\tilde{K}(t) = K'(0)t + \frac{1}{2}K''(0)t^2 + \dots + \frac{1}{k!}K^{(k)}t^k, \quad k \geq 2.$$

This is essentially the approach taken by Easton and Ronchetti (1986) for densities, with $k = 4$, which we return to later. Inference about the parameter $\mu = K'(0)$ is typically based on the statistic $V = n^\delta (\bar{X} - \mu)$, with $\delta > 0$. With $\delta = 1/2$, this is similar to the transformation that Feuerverger (1989) suggested when investigating the empirical saddlepoint for the density of a mean. We maintain the use of a general δ (although in practice we usually take $\delta = 1/2$). The reason multiplying the statistic by n^δ may not be obvious, but it turns out that this can result in the saddlepoints being of an order less than $O(1)$, which is needed in the proof of the following theorem.

We will denote the true distribution of V by $F_n(v)$, and the second order saddlepoint approximation, that is the first term of (??), by $G_n(v)$. The approximate saddlepoint approximation for the mean, resulting from using $\tilde{K}(t)$ in place of the true sampling *cgf* $K(t)$ in the usual saddlepoint formula, will be denoted by $\tilde{G}_n(v)$. In particular we are interested in knowing how well $\tilde{G}_n(v)$ approximates $F_n(v)$, given by (??).

Assumptions for the theory in this section are as follows:

- (1) $\tilde{K}^{(r)}(0) = K^{(r)}(0)$ for all $r \leq k$, $k \geq 2$;
- (2) $\tilde{K}^{(r)}(0) = 0$ for all $r > k$;
- (3) $\tilde{K}(0) = K(0) = 0$;
- (4) X_i , $i = 1, 2, \dots, n$ are independently and identically distributed.

Theorem 1 *Suppose $\tilde{K}(t)$ is an estimate of the *cgf* $K(t)$ and $\tilde{G}_n(v)$ approximates the true saddlepoint approximation $G_n(v)$ by using $\tilde{K}(t)$ in place of $K(t)$ in the saddlepoint formula. Under assumptions (1)-(4), for $\delta > 0$ we have*

$$(2) \quad \frac{\tilde{G}_n(v)}{G_n(v)} = 1 + O\left(n^{-\min\{(k-1)\delta, ((k+1)\delta-1)\}}\right).$$

Note also that if $\delta = \frac{1}{2}$, $(k-1)\delta = (k+1)\delta - 1$.

The proof of this theorem is given in Appendix A. Note that implicit in the theorem above is that a unique saddlepoint for $\tilde{K}(t)$ exists. This theorem leads to the following corollary that addresses the error of the approximate saddlepoint approximation of $n^\delta (\bar{X} - \mu)$ relative to its true distribution. Part of the proof of this next corollary involves demonstrating that none of the accuracy of the saddlepoint approximation is lost by transforming \bar{X} to $n^\delta (\bar{X} - \mu)$.

Corollary 1 *Suppose the distribution of $V = n^\delta (\bar{X} - \mu)$ is approximated using $\tilde{G}_n(v)$ as in the previous theorem. If the saddlepoint approximation to the mean attains an order of accuracy of $O(n^{-1})$, the relative error of $\tilde{G}_n(v)$ to the true distribution $F_n(v)$ is given by*

$$\frac{\tilde{G}_n(v)}{F_n(v)} = 1 + O\left(n^{-\min\{(k-1)\delta, ((k+1)\delta-1), 1\}}\right)$$

Note that in general, we take $\delta = 1/2$. For $\delta > 1/2$, the true distribution $F_n(v)$ does not converge and thus it only makes sense to consider $\delta \leq 1/2$. Furthermore, to obtain the greatest accuracy possible, we want to take δ as large as possible, making $\delta = 1/2$ the preferred choice. Upon doing so, if we take $k = 2$, that is, use a normal approximation, then the distribution $\tilde{G}_n(v)$ attains an accuracy of order $O(n^{-1/2})$, as expected. On the other hand, using the first four cumulants ($k = 4$) as in the Easton-Ronchetti approach (described in the next section), one obtains an accuracy of order $O(n^{-3/2})$. Thus, it is only necessary to use the first four cumulants so as to not lose any accuracy when approximating the cumulant generating function in this way.

3.1.1 The Easton-Ronchetti-Wang Approximation

We consider the approach of Easton and Ronchetti (1986), which was originally proposed for more general statistics than the sample mean, and the subsequent modification by Wang (1992). We begin with a review of Easton and Ronchetti's approach.

Any cumulant generating function may be expanded using a Taylor's series expansion. Thus $K(t)$ may be written

$$(3) \quad K(t) = \mu t + \frac{1}{2}\sigma^2 t^2 + \frac{1}{6}\rho_3\sigma^3 t^3 + \frac{1}{24}\rho_4\sigma^4 t^4 + \dots,$$

where,

$$\begin{aligned}\mu &= E(X), \\ \sigma^2 &= E(X^2) - E(X)^2, \\ \rho_r &= K_r/\sigma^r \text{ for } r \geq 3,\end{aligned}$$

and

$$K_r = \left(\frac{d}{dt} \right)^r K(t) \Big|_{t=0}$$

are the cumulants. The values $\rho_i, i \geq 3$ are the standardized cumulants, with ρ_3 being a measure of skewness (equal to zero for symmetric distributions) and ρ_4 a measure of kurtosis, or “long-tailedness”.

If X_1, \dots, X_n are iid, the cumulant generating function of \bar{X} is

$$(4) \quad K_{\bar{X}}(t) = \mu_V t + \frac{1}{2n} \sigma^2 t^2 + \frac{1}{6n^2} K_3 t^3 + \frac{1}{24n^3} K_4 t^4 + \dots$$

Easton and Ronchetti (1986) transform this to the quantity $R(t) = K_{\bar{X}}(nt)/n$ and propose approximating $R_V(t)$ with the first four terms of the Taylor series of $R(t)$ as determined from equation (??). We then have

$$\tilde{R}(t) = \tilde{K}_{\bar{X}}(nt)/n = \mu t + \frac{1}{2} \sigma^2 t^2 + \frac{1}{6} K_3 t^3 + \frac{1}{24} K_4 t^4.$$

This approximate cumulant generating function can replace the usual cumulant generating function in the saddlepoint equation to obtain an approximate saddlepoint approximation.

The previous section, which considers this particular approximation for the sample mean more generally, emphasizes that we should work with the quantity $n^{1/2}(\bar{X} - \mu)$ instead of \bar{X} directly, thus allowing us to infer about μ and maintain an order of accuracy of $O(n^{-1})$. If one could claim that the integrated saddlepoint distribution attained, in general, an order of $O(n^{-3/2})$ then the approximate saddlepoint distribution would maintain that same order of accuracy.

The Easton-Ronchetti truncated *cgf*, unfortunately, does not always yield an appropriate saddlepoint. In fact, it may not be possible to form an approximation at all. This is because there may not exist a real root of $K'(\hat{t}) = x$ with $\tilde{K}''(\hat{t}) > 0$. Due to this problem, Wang (1992) suggested modifying the Easton-Ronchetti approach to approximate the *cgf* by

a weighted average of a normal *cgf* (with mean and variance given by the true mean and variance) and the Easton-Ronchetti truncated *cgf*. The Wang modification is

$$(5) \quad \widetilde{K}(t; b) = \mu t + \frac{1}{2} \sigma^2 t^2 + g_b(t) \left(\frac{1}{6} K_3 t^3 + \frac{1}{24} K_4 t^4 \right),$$

where

$$g_b(t) = \exp \left\{ -\sigma^2 b^2 t^2 / 2n \right\}$$

and

$$b = \max \left\{ \frac{1}{2}, \inf \left\{ a : \widetilde{K}''(t; a) \geq 0 \text{ for all } t \right\} \right\}.$$

Since $g_b(t) = 1 + O(n^{-1})$, this approach will approximate the true with the same accuracy as with the Easton-Ronchetti procedure, when considered asymptotically. In practice, one might expect the approximation to perform better than the normal, but perhaps not as well as the original integrated saddlepoint approximation. In examples that we considered, b was typically very close to $1/2$ and so the value of $g_b(t)$ was in general quite small. Thus, we expect the accuracy to be closer to that expected from the Easton-Ronchetti approach.

3.2 General Theory - Estimation

The next theorem, similar that in the previous section, allows for situations where the *cgf* is *estimated* by considering observed data and constructing estimates of unknown parameters or the distribution function itself. In general, in the parametric setting, the cumulants will have the same order of accuracy as the accuracy of the estimated parameters.

We begin with the following assumptions on the *cgf*:

- (1') For some $\alpha > 0$, $\left. \frac{\partial^r}{\partial s^r} \tilde{K}(s) \right|_{s=t} = \left. \frac{\partial^r}{\partial s^r} K(s) \right|_{s=t} + O_p(n^{-\alpha})$, $\forall t \in \Re$ and $r = 0, 1, 2, \dots$
- (2') $\tilde{K}(0) = K(0) = 0$
- (3') $X_i, i = 1, 2, \dots, n$ are independently and identically distributed.

Again, we must examine the saddlepoint approximation for $n^\delta (\bar{X} - \tilde{K}'(0))$. By standardizing by $\tilde{K}'(0)$, we allow ourselves the possibility of letting $\tilde{K}'(0) = K'(0)$ and thus allowing for proper inference about $K'(0) = \mu$. The theorem follows.

Theorem 2 Let $G_n(v)$ be the true integrated saddlepoint approximation of the statistic $n^\delta (\bar{X} - \tilde{K}'(0))$ based on using the true cgf $K(t)$. Suppose $\tilde{G}_n(v)$ estimates $G_n(v)$ by using $\tilde{K}(t)$ as an estimate of $K(t)$ in the saddlepoint formula. Under assumptions (1')-(3'),

$$\frac{\tilde{G}_n(v)}{G_n(v)} = 1 + O_p \left(n^{-\min\{\alpha, (\alpha+2\delta-1)\}} \right).$$

Note that when $\delta = 1/2$, $\alpha + 2\delta - 1 = \alpha$.

The proof of this theorem is similar to that of Theorem 1. Also, this result here is similar to, but much more general than what Feuerverger obtained for the empirical saddlepoint density, where he only considered $\alpha = \delta = 1/2$ and $\tilde{K}'(0) = \bar{x}$, the observed mean.

Again, the estimated, integrated saddlepoint approximation can be compared to the true distribution via the following corollary.

Corollary 2 Denote the true distribution of $V = n^\delta (\bar{X} - \mu)$ by $F_n(v)$. Under the conditions of Theorem ??, if

$$\frac{G_n(v)}{F_n(v)} = 1 + O(n^{-\eta})$$

with $\eta \geq 1$, then

$$\frac{\tilde{G}_n(v)}{F_n(v)} = 1 + O_p \left(n^{-\min\{\alpha, (\alpha+2\delta-1), \eta\}} \right).$$

Next we consider the implications of this theory on specific methods of estimation of the cumulant generating function. For example, the form of the density (and the cgf) may be known with the exception of the unknown parameter values. In such a case we may find the maximum likelihood estimates for these parameters and substitute them into the saddlepoint formula. Again, this typically results in a cgf which estimates the true cgf to an order of $O_p(n^{-1/2})$. We have the following corollary.

Corollary 3 Suppose the sampling cumulant generating function of $V = n^\delta (\bar{X} - \mu)$ is known except for the values of unknown parameter, and these parameters are estimated with \sqrt{n} -consistent estimators and substituted into the cgf.

Denote the resulting estimated saddlepoint approximation by $\hat{G}_n(v)$. Then

$$\frac{\hat{G}_n(v)}{F_n(v)} = 1 + O_p \left(n^{-\min\{\alpha, (\alpha+2\delta-1), 1/2\}} \right)$$

Typically, the unknown parameters will be estimated using maximum likelihood, which often results in \sqrt{n} -consistent estimators; see Lehmann (1983, Section 6.3).

In contrast to the above setting, one may be unwilling to assume anything about the form of the *cgf* and wish to take a completely nonparametric approach. One of the papers to introduce the idea of using the empirical *cgf* in the saddlepoint was written by Davison and Hinkley (1988), but this idea was also used by Robinson (1982). The following corollary addresses this use of the empirical cumulant generating function.

Corollary 4 *Denote the true distribution $\hat{V} = n^{1/2}(\bar{X} - \bar{x})$ by $F_{\hat{V}}$, and suppose that its sampling cumulant generating function is estimated by the empirical *cgf*. Denote the resulting estimated saddlepoint approximation by $\hat{G}_{\hat{V}}$. Then*

$$\frac{\hat{G}_{\hat{V}}(v)}{F_{\hat{V}}(v)} = 1 + O_p(n^{-1/2}).$$

This corollary parallels the theorem of Feuerverger (1989). Unfortunately, the conclusion of this corollary does not help us in making inferential statements about the mean μ . However, using a combination of the empirical moments and the Easton-Ronchetti approach, inference about the mean is possible. In particular, we could assume that the true mean is some value μ_0 and estimate higher order cumulants using the empirical cumulants. Thus, one would estimate the *cgf* by

$$\tilde{K}(t) = \mu_0 t + \frac{1}{2} \tilde{\sigma}^2 t^2 + \frac{1}{3!} \tilde{K}_3 t^3 + \frac{1}{4!} \tilde{K}_4 t^4$$

where

$$\begin{aligned} \tilde{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \\ \tilde{K}_3 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^3 \\ \tilde{K}_4 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^4 - 3\tilde{\sigma}^4 \\ \tilde{K}_i &= 0 \text{ for } i \geq 5. \end{aligned}$$

By the central limit theorem, these cumulants are estimated with accuracy $O_p(n^{-1/2})$. Then, by considering the recentered, integrated saddlepoint

approximation, we can obtain a saddlepoint approximation which is accurate to $O_p(n^{-1/2})$. The theorem that follows addresses this situation.

First, we modify the assumptions of the previous theorem. In particular, we now assume

(1'') For $\alpha > 0$, $\tilde{K}^{(r)}(0) = K^{(r)}(0) + O(n^{-\alpha})$ for all $r \leq k$;

(2'') $\tilde{K}^{(r)}(0) = 0$ for all $r > k$;

(3'') $\tilde{K}(0) = K(0) = 0$;

(4'') $X_i, i = 1, 2, \dots, n$ are independently identically distributed.

Theorem 3 *Suppose $\tilde{K}(t)$ is an estimate of the cgf $K(t)$ and $\tilde{G}_n(v)$ approximates the true saddlepoint approximation $G_n(v)$ by using $\tilde{K}(t)$ in place of $K(t)$ in the saddlepoint formula. Under assumptions (1'')-(4''), for $\delta > 0$ we have*

$$\frac{\tilde{G}_n(v)}{G_n(v)} = 1 + O(n^{-\xi})$$

where $\xi = \min\{\alpha, \alpha + 2\delta - 1, k\delta, (k+1)\delta - 1\}$.

In the case that $\delta = \frac{1}{2}$,

$$\xi = \min\{\alpha, (k-1)/2\},$$

so it seems that this latter method of estimating the sample mean distribution using the empirical cgf is no worse than the mle saddlepoint (estimating the cgf using parameter mle's). However, this might not be quite true. The mle saddlepoint is essentially imitating the parametric bootstrap. The difference is that the additional error, other than the error due to the maximum likelihood estimation, comes not through the number of resamples, but through analytical calculations.

In addition, with respect to the order of convergence, the mle saddlepoint and the empirical saddlepoint perform asymptotically no better than the normal approximation. We know that the bootstrap does have some properties which suggest that it is an improvement to the nonparametric approach. In particular, the bootstrap is asymptotically minimax and has the smallest asymptotic mean square error (See Shao and Tu, 1995).

Lastly, the orders of convergence carry through to approximating the true distribution function.

Corollary 5 Denote the true distribution of $V = n^\delta (\bar{X} - \mu)$ by $F_n(v)$. Under the conditions of Theorem ??, if

$$\frac{G_n(v)}{F_n(v)} = 1 + O(n^{-1}),$$

then

$$\frac{\widetilde{G}_n(v)}{F_n(v)} = 1 + O_p(n^{-\min\{\xi, 1\}}).$$

4 Example

An empirical study of the various methods might enhance our understanding of their comparative accuracy. As the theory is asymptotic in nature, it may not give the full picture for small sample sizes.

We consider random variables, X_i , drawn from a Weibull distribution. Inference concerning the scale parameter β is available by studying the values of $y_i = x_i/\beta$. The density is given by

$$f(y_i|\gamma) = \gamma y_i^{\gamma-1} e^{-y_i^\gamma}.$$

If γ were known, \bar{Y} would be pivotal for beta. For inference about β , let $y_{\alpha,\gamma}$ represent the cut-off points, where the distribution function satisfies $F(y_{\alpha,\gamma}; \gamma) = \alpha$. To find these cut-off points, we use the saddlepoint approximation to the distribution of \bar{Y} evaluated at the mle of γ . A confidence interval for β is then given by

$$\left(\frac{\bar{x}}{y_{\alpha/2;\hat{\gamma}}}, \frac{\bar{x}}{y_{1-\alpha/2;\hat{\gamma}}} \right).$$

Bain and Engelhardt (1981) provide a simple procedure for finding confidence limits for the scale parameter. These are based on the distribution of the pivotal quantity $(n-1)^{1/2} \hat{\gamma} \ln(\hat{\beta}/\beta) / 1.053$, which asymptotically has a t -distribution with $n-1$ degrees of freedom. Hence, a $(1-\alpha) \times 100\%$ confidence interval is given by

$$\exp \left(\log(\beta) \pm t(1-\alpha/2; n-1) \times 1.053 \times \frac{1}{\hat{\gamma} \sqrt{n-1}} \right).$$

Theoretically, by using the mle of γ in the saddlepoint approximation, the accuracy of the former interval will be reduced down at least to the level

of accuracy of the latter. Table ?? gives results of simulations designed to study the small sample properties of these intervals for just one set of true parameter values, $\gamma = 3$ and $\beta = 4$, sample size $n = 10$, and confidence level 90%. The number of simulations was 1000. The saddlepoint approach could

Table 1: Simulation study comparing the saddlepoint and the Bain and Engelhardt (1981) t-interval; $\gamma = 3$, $\beta = 4$, $n=10$, confidence level 0.90., and 1,000 simulations.

	Avg. # of times CI overshoot	Avg. # of times CI missed	Average length of CI	St. Dev. of CI length
Saddlepoint CI	0.058	0.130	1.414	0.114
Asymptotic t-interval	0.042	0.101	1.557	0.159

be improved upon further by noting that in the interval above, $y_{\alpha/2;\hat{\gamma}}$ and $y_{1-\alpha/2;\hat{\gamma}}$ may be replaced by $y_{\alpha_1;\hat{\gamma}}$ and $y_{1-\alpha_2;\hat{\gamma}}$, respectively where $\alpha_1 < 0.50$, $\alpha_1 + \alpha_2 = \alpha$, and the distance between $1/y_{\alpha_2;\hat{\gamma}}$ and $1/y_{\alpha_1;\hat{\gamma}}$ is minimized. This will create the shortest interval of this kind at a significance level of α .

Note that these confidence intervals, even if the distribution of \bar{Y} were known analytically, would not be exact due to estimation of the parameter γ . This problem can be circumvented by calibrating the interval. See, for example, Loh (1987) and Hall (1992). Alternatively, one could, for each value of β considered, find the associated value of γ . This approach mimics a profile likelihood approach. One drawback of this approach is that for small values of γ , the confidence interval does not exist, due to the non-existence of the cumulant generating function for values of γ which are less than 1.0.

5 Discussion

We have seen, both analytically and empirically, that improvements obtained by the saddlepoint methodology may not be large when using estimates for the nuisance parameters. As suggested in the numerical example, use of bootstrap calibration may improve the accuracy of inferences made using these saddlepoint approximations.

On the other hand, note that saddlepoint approximations may be used to simplify bootstrap calculations in that they provide an analytical approximation where extensive re-sampling might otherwise be needed. It has been well-established (see for example Hall, 1992) that bootstrapping the studentized mean, $(\bar{X} - \mu)/S$, often results in more accurate inferences that

are $O_p(n^{-1})$ instead of $O_p(n^{-1/2})$. Consequently, it is expected that the saddlepoint approximation would gain a similar increase in accuracy when considering the studentized mean.

The saddlepoint approximation to the density of the studentized mean has already been examined in the nonparametric case by Daniels and Young (1991). Their approximation is based on a transformation and marginalization of the joint distribution of \bar{X} and the sample mean of X^2 . Ohman (1997) considers this approach in both the nonparametric and the parametric setting. When the original distribution is the normal, the saddlepoint approximation is exact, yielding the t -distribution. In most other cases, however, the joint cumulant generating function does not exist. (We also note that even if the joint *cgf* of \bar{X} and S existed, this quantity may not be the best quantity for inference about the mean. Although the studentized mean will, in most cases, be asymptotically pivotal, dividing by the sample standard deviation, S , does not always produce a pivot.)

Unfortunately, the accuracy of the estimated and approximated saddlepoint approximation, in the studentized mean case, has not been thoroughly assessed, and may require tools beyond those developed here. We outline two alternative approaches, which are left for future research, with the goal of improving the accuracy of the saddlepoint approximations to order $O_p(n^{-1})$ from $O_p(n^{-1/2})$ when additional parameters in the approximation have been estimated to order $O_p(n^{-1/2})$.

Upon reviewing the conjugate density (exponential tilting) derivation of the saddlepoint approximation for densities, one sees that the cumulant generating function becomes central due to the formulation of the conjugate density. If we can find a function other than the exponential function, which allows simple calculation of the first four moments of the conjugate density, we can apply a modified saddlepoint formula, using an Easton-Ronchetti/Wang approach, even if the cumulant generating function itself does not exist. One implementation for the saddlepoint approximation to the density of the studentized mean, using a polynomial approximation, is developed in Ohman (1997).

A second alternative to approximating the distribution of the studentized mean may be the use of mixture distributions. For example, the t -distribution has no cumulant generating function. However, it may be written as a mixture of normals, for which cumulant generating functions do exist. Thus, one could truncate the mixture distribution, find the corresponding cumulant generating function (probably numerically), and use

this “approximate cgf” in place of the non-existent one in the saddlepoint approximation. Careful attention needs to be given to how truncation of the mixture distribution affects the accuracy of the saddlepoint approximation.

As a final note on pivotal quantities, one possible approach is to approximate the *cgf* of the pivotal statistic itself in a manner similar to that demonstrated by Hall (1992, §2.3), that is by using an asymptotic expansion. One might then use a this approximate *cgf* in the saddlepoint approximation directly. It is expected that this would produce a highly accurate approximations as it does in the case of the bootstrap. However, due to the form of the cumulant generating function, it is unclear how to prove asymptotic equivalency to the true distribution. Techniques used in this paper are not sufficient because of the need to assign a rate of at least $O(n^{-1/2})$ to the saddlepoints themselves.

Appendix A: Proof of Theorem 1

The ratio of the second order saddlepoint approximations is given by:

$$(6) \quad \frac{\exp \left\{ n \left(\tilde{K}(\tilde{t}) - n^{-\delta} \tilde{t}v - \tilde{t}\tilde{\mu} \right) \right\}}{\exp \left\{ n \left(K(t) - n^{-\delta} tv - t\mu \right) \right\}} \times \frac{\sqrt{K''(t)}}{\sqrt{\tilde{K}''(\tilde{t})}} \times \frac{t}{\tilde{t}}$$

We first examine the first fraction of this quantity, the fraction that becomes

$$\exp \left\{ n \left(\left(\tilde{K}(\tilde{t}) - \tilde{t}\mu \right) - \left(K(t) - t\mu \right) - n^{-\delta}(\tilde{t} - t)v \right) \right\}$$

To determine the orders of t , \tilde{t} and $\tilde{t} - t$, note that the equation that is solved to find the “saddlepoint” can be written

$$n^{-\delta}v = K'(t) - K'(0) = K''(0)t + \frac{1}{2}K'''(0)t^2 + \dots,$$

where $K'(0) = \mu$. Thus, upon solving this equation for t , we see that the saddlepoint $t = O(n^{-\delta})$ and $\tilde{t} = O(n^{-\delta})$.

Next, write

$$\begin{aligned} 0 &= n^{-\delta}(v - v) \\ &= \left(\tilde{K}'(\tilde{t}) - \tilde{\mu} \right) - \left(K'(t) - \mu \right) \\ &= \left[\tilde{K}''(0)\tilde{t} - K''(0)t \right] + \frac{1}{2} \left[\tilde{K}'''(0)\tilde{t}^2 - K'''(0)t^2 \right] + \dots, \end{aligned}$$

and note that in the case that $k = 2$, all approximated cumulants of order three and above are zero, so the above quantity is equivalent to

$$K''(0) (\tilde{t} - t) - \frac{1}{2} K'''(0) t^2 - \dots$$

If $\tilde{t} - t = O(n^{-\zeta})$, the above implies that $0 = O(n^{-\zeta}) + O(n^{-2\delta})$, and thus when $k = 2$, $\zeta = 2\delta$. If $k = 3$,

$$\begin{aligned} 0 &= K''(0) (\tilde{t} - t) + \frac{1}{2} K'''(0) (\tilde{t} - t) (\tilde{t} + t) - \frac{1}{3!} K^{(4)} t^3 - \dots \\ &= O(n^{-\zeta}) + O(n^{-(\zeta+\delta)}) + O(n^{-3\delta}), \end{aligned}$$

which implies $\tilde{t} - t = O(n^{-3\delta})$. Continuing, for general k we have that $\tilde{t} - t = O(n^{-k\delta})$.

Now, return to the ratio (??), and consider the quantity

$$(7) \quad \left(\tilde{K}(\tilde{t}) - \tilde{t}\tilde{\mu} \right) - (K(t) - t\mu)$$

in the exponent. By expanding each of $K(t)$ and $\tilde{K}(\tilde{t})$ in a Taylor series and subtracting, we find that (??) is equivalent to

$$\frac{1}{2} \left[\tilde{K}''(0) \tilde{t}^2 - K''(0) t^2 \right] + \frac{1}{3!} \left[\tilde{K}'''(0) \tilde{t}^3 - K'''(0) t^3 \right] + \dots$$

If $k = 2$, (??) equals

$$\frac{1}{2} K''(0) (\tilde{t} - t) (\tilde{t} + t) - \frac{1}{3!} K'''(0) t^3 - \dots,$$

which is $O(n^{-3\delta})$. For the case that $k = 3$, (??) equals

$$\frac{1}{2} K''(0) (\tilde{t} - t) (\tilde{t} + t) + \frac{1}{3!} K'''(0) t^3 (\tilde{t} - t) (\tilde{t}^2 + \tilde{t}t + t^2) - \dots,$$

which is $O(n^{-4\delta})$. For general k , this portion of the exponent (??) is $O(n^{-(k+1)\delta})$. Moreover, we also have that $n^{-\delta} v (\tilde{t} - t) = O(n^{-(k+1)\delta})$, and putting this altogether yields

$$(8) \quad \tilde{K}(\tilde{t}) - K(t) - (\tilde{t} - t)\mu - n^{-1/2}(\tilde{t} - t)v = O(n^{-(k+1)\delta}).$$

Multiplying by n to find that the complete exponent then gives $O\left(n^{-((k+1)\delta-1)}\right)$, and finally for the ratio, by considering a Taylor series of the exponential, we have that $\exp\{\} = 1 + O\left(n^{-((k+1)\delta-1)}\right)$.

We now turn our attention to the second part of the fraction, $\sqrt{K''(t)/\tilde{K}''(\tilde{t})}$ given in (??), dividing it into two parts. We examine the ratio of second derivatives by first considering each individually. Subtracting Taylor series yields

$$\begin{aligned}\tilde{K}''(\tilde{t}) - K''(t) &= \left(\tilde{K}''(0) - K''(0)\right) + \left(\tilde{K}'''(0)\tilde{t} - K'''(0)t\right) \\ &\quad + \frac{1}{2}\left(\tilde{K}''''(0)\tilde{t}^2 - K''''(0)t^2\right) - \dots\end{aligned}$$

For $k = 2$, this equals

$$-K'''(0)t - \frac{1}{2}K''''(0)t^2 - \dots = O\left(n^{-\delta}\right),$$

for $k = 3$, this equals

$$K'''(0)(\tilde{t} - t) - \frac{1}{2}K''''(0)t^2 - \dots = O\left(n^{-2\delta}\right)$$

continuing, for any k , $\tilde{K}''(\tilde{t}) - K''(t) = O\left(n^{-(k-1)\delta}\right)$. It then follows that,

$$\begin{aligned}\left[\frac{K''(t)}{\tilde{K}''(\tilde{t})}\right]^{-1/2} &= \left[\frac{K''(t)}{K''(t) + O\left(n^{-(k-1)\delta}\right)}\right]^{-1/2} \\ &= 1 + O\left(n^{-(k-1)\delta}\right).\end{aligned}$$

Finally, we show that the ratio of saddlepoints $t/\tilde{t} = 1 + O\left(n^{-(k-1)\delta}\right)$. Begin by observing that since

$$K'(t) = n^{-\delta}v + \mu$$

and

$$K'(t) = K'(0) + K''(0)t + \frac{1}{2}K'''(0)t^2 + \dots,$$

where $K'(0) = \mu$, we can write

$$K''(0)t = n^{-\delta}v - \frac{1}{2}K'''(0)t^2 - \frac{1}{3!}K^{iv}(0)t^3 - \dots$$

Thus, in considering the ratio of saddlepoints,

$$\begin{aligned}
\frac{t}{\tilde{t}} &= \frac{(tK''(0)) / (K''(0))}{(\tilde{t}\tilde{K}''(0)) / (\tilde{K}''(0))} \times \frac{\tilde{K}''(0)}{K''(0)} \\
&= \frac{n^{-\delta}v - \frac{1}{2}K'''(0)t^2 - \dots}{n^{-\delta}v - \frac{1}{2}\tilde{K}'''(0)\tilde{t}^2 - \dots} \\
&= \frac{v - \frac{1}{2}K'''(0)n^{\delta}t^2 - \dots}{v - \frac{1}{2}\tilde{K}'''(0)n^{\delta}\tilde{t}^2 - \dots}.
\end{aligned}$$

Note that for $k \geq 2$, $\tilde{K}''(0) = K''(0)$. In particular, upon recalling that $t = O(n^{-\delta})$, this becomes, for $k = 2$,

$$\frac{v - \frac{1}{2}K'''(0)n^{\delta}t^2 - \dots}{v} = 1 + O(n^{-\delta}) ;$$

for $k = 3$,

$$\frac{v - \frac{1}{2}K'''(0)n^{\delta}t^2 - \frac{1}{3!}K^{(iv)}(0)n^{\delta}t^3 - \dots}{v - \frac{1}{2}\tilde{K}'''(0)n^{\delta}\tilde{t}^2} = 1 + O(n^{-\delta}) ;$$

and so on. Therefore, in general,

$$\begin{aligned}
\frac{t}{\tilde{t}} &= \frac{tK''(0)}{\tilde{t}\tilde{K}''(0)} \\
&= 1 + O(n^{-(k-1)\delta}).
\end{aligned}$$

Multiplying these three parts together then, we get that the ratio is

$$(1 + O(n^{-((k+1)\delta-1)})) (1 + O(n^{-(k-1)\delta})) (1 + O(n^{-(k-1)\delta})) = 1 + O(n^{-\min\{(k+1)\delta-1, (k-1)\delta\}})$$

■

References

- [1] Blæsild P. and Jensen, J.L. (1985). Saddlepoint formulas for reproductive exponential models. *Scandinavian Journal of Statistics*, **12**, 193-202.
- [2] Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631-650.

- [3] Daniels, H. E. (1980). Exact saddlepoint approximations. *Biometrika* **67**, 59-63.
- [4] Daniels, H. E. (1983). Saddlepoint approximations for estimating equations. *Biometrika* **70**, 89-96.
- [5] Daniels, H. E. (1987). Tail Probability Approximations. *Int. Statist. Rev.* **55**, 37-48.
- [6] Daniels, H.E. and Young, G.A. (1991). Saddlepoint approximation for the studentized mean with an application to the bootstrap. *Biometrika*. **78**, 169-179.
- [7] Davison, A.C. and Hinkley, D.V. (1988). Saddlepoint approximations in resampling methods. *Biometrika*, **75**, 417-431.
- [8] Easton, G.S. and Ronchetti, E. (1986). General saddlepoint approximations with applications to L statistics. *Journal of the American Statistical Association*, **81**, 420-430.
- [9] Feuerverger, A. (1989). On the empirical saddlepoint approximation. *Biometrika*, **76**, 457-464.
- [10] Field, C. and Ronchetti, E. (1990). *Small Sample Asymptotics*. Hayward, CA: Institute of Mathematical Statistics Lecture Notes.
- [11] Gatto, R. and Ronchetti, E. (1996). General saddlepoint approximations of marginal densities and tail probabilities. *Journal of the American Statistical Association*, **91**, 666-673.
- [12] Goutis, C. and Casella, G. (1995). Explaining the saddlepoint approximation. Technical Report BU-1311-M. Biometrics Unit, Cornell University, Ithaca, NY.
- [13] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag: New York.
- [14] Jensen, J.L. (1995). *Saddlepoint Approximations*, Clarendon Press: Oxford.
- [15] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion* New York: Springer-Verlag.

- [16] Kolassa, J. E. (1994). *Series Approximation Methods in Statistics*. New York: Springer-Verlag.
- [17] Lehmann, E. L. (1983). *Theory of Point Estimation*. New York: Springer-Verlag.
- [18] Loh, W.Y. (1987). Calibrating confidence coefficients. *Journal of the American Statistical Association*, **82**, 155-162.
- [19] Ohman, P. A. (1997). Estimated and Approximated Saddlepoint Approximations. PhD. Thesis, Department of Statistical Science, Cornell University.
- [20] Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.* **3**, 213-238.
- [21] Reid, N. (1991). Approximations and asymptotics. *Statist. Theory Model., Essays in Honor of D. R. Cox*. London: Chapman and Hall, 287-334.
- [22] Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *Journal of the Royal Statistical Society, B*, **44**, 91-101.
- [23] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer-Verlag: New York.
- [24] Wang, S. (1992). General saddlepoint approximations in the bootstrap. *Statistics & Probability Letters*, **13**, 61-66.